Title: Analyzing Health Effects of Chemical Exposure using R Markdown

Description:

We are exposed to numerous environmental chemicals each day. We are interested in quantifying the health effects of environmental chemical mixtures, assessing joint actions, and identifying the interactions of combined chemicals. Analyzing health effects of chemical exposures can contribute to preventive measures to mitigate the potential impact of these exposures.

In this project, we aim to summarize advanced statistical approaches for analysis of complex mixtures and knit them to R tutorials to make them accessible to researchers without extensive statistics or mathematics backgrounds. This will include online tutorials to introduce advanced statistical approaches to scientists and to provide examples of their use using national survey data. In 2015, the National Institute of Environmental Health Sciences (NIEHS) convened a workshop to bring together experts to "identify and compare statistical approaches for analyzing chemical mixture data in epidemiological studies". The collection of abstract, code, and datasets from the workshop are all available on the NIEHS website (https://www.niehs.nih.gov/news/events/pastmtg/2015/statistical/index.cfm). NIEHS also launched a funding initiative to address the analytical challenges of environmental mixtures research, called Powering Research through Innovative methods for Mixtures in Epidemiology (PRIME) program. Various new statistical methods supported by the NIEHS PRIME Program are available in https://github.com/niehs-prime/.

In this project, you will learn advanced statistical approaches on analyzing chemical mixtures, including classic linear regression, classification and prediction methods (such as principal component analysis (PCA), sparse PCA), exposure-response surface estimation (such as exposure surface smoothing, Bayesian networks), and variable selection and shrinkage strategies (such as LASSO, weighted quantile sum regressions). You will gain a working knowledge of these methods and be able to communicate them with people without extensive statistics or mathematics background. You should be familiar with the R language so that you will be able to apply these methods to answer scientific questions using R. You should be familiar with R markdown to write reproducible reports and able to use GitHub to collaborate with other researchers.

By the end of the project, we will have an online tutorial (website) introducing statistical approaches for chemical mixtures analysis and applications using R. For students interested in research, at the end of this project, we may consider a new research project focusing on developing statistical models and methods for environmental chemical mixtures. A possible side project, depending on skills and interests, includes improving workflow using virtual lab notebooks for data collection and annotation.

Project categories: educational, research

Students' year: sophomore, junior, senior

Student qualification: STA 198/199 or the equivalent (skills with R, R Markdown, and GitHub)

Hours per week: 5-7, 8-10. Students are expected to attend regular weekly meetings (1 hour), and work on the project for 6-8 hours/week

Compensation:

Paid-general / paid-work required.

$15/hr. This project is funded by the Superfund Research Center at Duke University.

Number of openings: (?)

Contact Information:

Mentor: Yunran Chen (yc304@duke.edu)

Supervisor: Prof. Amy Herring (amy.herring@duke.edu)